DOCUMENT RESUME

ED 310 362 CS 009 766

AUTHOR Seda, Ileana
TITLE Assessment Format and Comprehension Performance.
PUB DATE 89
NOTE 36p.; Paper presented at the Annual Meeting of the International Reading Association (34th, New Orleans, LA, April 30-May 4, 1989).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Comparative Analysis; Elementary Education; *Evaluation Methods; Grade 3; Grade 6; *Multiple Choice Tests; Protocol Analysis; *Reader Response; *Reading Comprehension; Reading Research; *Reading Tests; Test Format; Urban Schools
IDENTIFIERS Illinois; Illinois (Chicago); Text Factors

ABSTRACT

In response to the tradition of examining a new test's validity by comparing it with a well-established multiple-choice test, a study investigated whether multiple-choice tests with one right answer can measure the reading comprehension process as defined by constructivists. The study compared the information obtained from two different multiple-choice tests (one traditional with one good answer per question and one novel with multiple good answers per question) and from two free-response measures (retellings and semi-structured interviews). Subjects were 28 third graders and 25 sixth graders in Illinois--two schools were in the inner city of Chicago, one was suburban, and two were in central Illinois. Of particular interest were readers' text understandings and readers' rankings on the four measures. Qualitative results (from case studies) indicated that subjects' performance varied according to the assessments used. Results also pointed to discrepancies in the use of text characteristics and salient text information by subjects and by test questions. Different assessment formats, then, seem to tap different processes or different components of one process. The study raises questions about task demands, text characteristics, reader factors, and sources of responses in reading assessment. (One figure is included and 30 references are attached.) (SR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED310362

ASSESSMENT FORMAT AND COMPREHENSION PERFORMANCE

Ileana Seda

The Pennsylvania State University

253 Chambers Building

University Park, PA 16802

CS009766

BEST COPY AVAILABLE

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Ileana Seda

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

ASSESSMENT FORMAT AND COMPREHENSION PERFORMANCE

(Abstract)

The literature in reading comprehension points to a disparity between current knowledge about the reading process and current assessment practices. This study was designed to compare the information obtained from two different multiple-choice tests (one traditional with one good answer per question and one novel with multiple good answers per question) and from two free-response measures (retellings and semi-structured interviews). Of particular interest were readers' text understandings and reader's rankings on the four measures. Qualitative (case studies) results indicated that subjects' performance varied according to the assessments used. Results also pointed to discrepancies in the use of text characteristics and salient text information by subjects and by test questions. Different assessments formats then, seem to tap different processes or different components of one process. The study raises questions about task demands, text characteristics, reader factors, and source of responses in reading assessment.

ASSESSMENT FORMAT AND COMPREHENSION PERFORMANCE

Prevalent theoretical views on reading comprehension favor a constructivist cognitive framework in which text meanings are determined by readers' interpretations. In this, view the reader takes an active role in constructing worlds of meaning from the clues left by the author on the printed page (Anderson, 1977; Anderson & Ortony, 1975; Bartlett, 1932; Bock & Brewer, 1974; Johnston, 1983; Norman & Rumelhart, 1975; Rumelhart & Ortony, 1977; Thorndike, 1917). The expectation is to have multiple text interpretations which will also vary qualitatively from reader to reader in terms of depth and sophistication of understanding.

In contrast, current assessment practices maintain single, pre-specified correct text interpretations with literal level understanding as their primary focus. Such practices raise questions about the use of test scores for making instructional decisions which intend to promote success in reading (Johnston, 1983, 1984b; Messick, 1975, 1981; Tierney & McGinley, 1987). Nevertheless, test scores are presently used as standards to measure general educational outcomes and individual educational success (Cole, 1984). Scores of reading tests are used to measure and account for schools' effectiveness, teachers' effectiveness, and placement of individuals into groups (Cole, 1984; Madaus, 1983). Johnston (1983) points to the discrepancy in cognitive demands of multiple-choice tests of reading comprehension, where the process is one of recognition versus

those of the reading comprehension process where the process is inferential.

Amidst controversy, the standards by which instructional effectiveness is measured have remained virtually unchanged (Farr & Carey, 1986). Presently, some states such as Michigan (Wixson, Peters, Weber, & Roeber, 1987; and Illinois (Valencia & Pearson, 1987) have developed large-scale assessments which attempt to bring reading theory and assessment practices closer together. An additional goal of these attempts is to provide a closer match between assessment and desirable instructional practices.

Meanwhile, alternatives to traditional reading tests have been and are being used extensively in reading research. Taylor (1984) has pointed out that free-recalls, retellings, and interview procedures have been increasingly appearing in the reading research literature. The trend to use free-response measures to assess performance presents a dichotomy in the operating standards at work in research and practice. Free-response measures in general have been rejected by the test and measurement tradition on grounds of inefficiency, lack of objectivity, and difficulty in quantification (Frederiksen, 1984).

Current practices in reading assessment have followed a psychometric model with emphasis on efficiency and inadequate concerns for validity. The tradition in psychometrics is to examine a new test's validity by comparing it with a well established multiple-choice test. The validity question here is: Can multiple-choice

tests with one right answer measure the reading comprehension process as defined by constructivists?

A reasonable step in the direction of addressing validity issues in testing is to compare the kinds of information that free-response and multiple-choice measures yield in terms of readers' understandings of texts and in terms of the ratings and rankings readers receive in each measure.

The present study addressed the validity issue by contrasting readers' comprehension performace on the following: (1) passages of a traditional multiple-choice test (short passages, one right answer), (2) passages of an experimental multiple-choice test (long passages, multiple good answers), (3) verbal retellings on the short (traditional) and long (experimental) test passages, and (4) verbal responses to semi-structured interviews on the same short and long test passages. The two sets of passages were administered to third and sixth graders who participated on a large-scale assessment project, the Reading Assessment Initiative in the State of Illinois.

Method

Subjects

Thirty subjects per grade level, six per school, comprised the initial sample of the study. The final sample consisted of 28 third and 25 sixth graders due to absences from school. Three schools per grade level were located in the Chicago metropolitan area since two thirds of the student population in Illinois is in that area. Two of these three schools were inner city (one middle and one low socio-

economic level) and one was suburban (lower end of mid socio-economic level). The remaining two schools were from central Illinois districts (mid socio-economic level for the disctrict). Subjects were average, at grade level, readers from a pool of twenty participants per school in the Reading Assessment Initiatives in the State of Illinois project.

Measures

Six assessment measures of reading comprehensions, three per passage, were administered to the subject groups:

(1) two traditional test passages (short) and corresponding multiple-choice items -- one right answer,

(2) two experimental test passages (long) and corresponding multiple-choice items -- multiple good answers,

(3) verbal retellings on the traditional short passages,

(4) verbal retellings on the experimental long passages,

(5) semi-structured interviews on the traditional passages,

(6) semi-structured interviews on the experimental passages.

The traditional multiple-choice test consisted of a narrative and an expository passage (200 - 300 words) and their corresponding questions (5 - 6 questions) from the Metropolitan Achievement Test (Prescott, Balow, Hogan, & Farr, 1984) included in the Illinois Inventory of Educational Progress, Inventory: 1 (Illinois State Board of Education & The Psychological Corporation, 1986a; 1986c). The experimental multiple-choice test consisted of a full-length narrative and exposition (1200 to 1500 words) each for third and sixth grade from

the Illinois Inventory of Educational Progress, Inventory: 3 (Illinois State Board of Education & The Psychological Corporation, 1986b; 1986d). Corresponding multiple-choice questions for those passages consisted of a more than one good answer format. In this format subjects were to select all the good answers to the questions (12 to 15 questions) from a total of 4 to 6 choices per question.

Semi-structured interview questions for each passage were written by the researcher based on text map criteria of important ideas (Warren, Nicholas and Trabasso, 1979; Schallert, Ullerick, and Tierney, 1984). Developing questions in this manner allowed the interviews to focus on significant textual information and avoid insignificant details.

Procedure

Test booklets and test results data for the subjects in this study were made available through the Reading Assessment Initiatives in the State of Illinois project. Free-response data for the 28 third graders and the 25 sixth graders were collected at each school site three to four weeks after the tests for the state project had been administered. Subjects and interviewers met for two consecutive days in dyadic (one on one) sessions of approximately 45 minutes each. Subjects performed the following tasks during each interview session:

1. Read the experimental passage (actually re-read it since they had read the passage during the state testing).
2. Retell the experimental passage.

3. Answer semi-structured interview questions for the experimental passage.
4. Read the traditional passage (re-read).
5. Retell the traditional passage.
6. Answer semi-structured interview questions for the traditional passage.

The first day the procedure was followed for the two narrative texts, the second day for the expository texts. All sessions were audio-taped. Subjects' first task was to retell the experimental text as prompted by the following: "Suppose you are going to tell into a tape recorder the story/article that you just read, so that a second/fifth grader who has not read it can understand even the hard parts of it by listening to your tape. Go ahead and tell the story/article." Retellings were unprompted, except for an initial prompt, and an occasional "M-hm," "Tell me more," or "Is there anything else?" The task was presented as a hypothetical situation of telling the passage to a second (third grade subjects) or fifth grader (sixth grade subjects) in order to lend some reality to it. It was hoped that the hypothetical situation would promote richer retellings than telling it for the interviewer who obviously had read the passage. Also, making the hypothetical situation for a student a year younger than themselves hopefully encouraged subjects to tell the passage as clearly as possible and minimized task threat. The same procedure was followed for the experimental text. This particular sequence allowed subjects to move from the most familiar to the less

familiar; from the traditional to the experimental test passages and from the narrative to the expository test passages.

The semi-structured interviews consisted of a series of questions based on text information. Interviewers asked a set of open-ended questions which allowed specific information about text understandings to surface even when subjects had not spontaneously provided it in their retellings. The intent was to allow subjects to elaborate on their original retellings and provide them opportunities to tell information they may have deliberately or inadvertently left out. Each question was followed by prompts as necessary in order to achieve a common understanding between interviewer and interviewee. These prompts were intended to be non-leading in nature, and took the form of "What do you mean?," " Tell me more about that," and repetitions in question form of subjects' assertions.

Scoring and Data Analysis

Case study analyses were conducted to trace and compare individual cases' performance on the measures used. Total interview protocols (retellings and semi-structured interviews) were transcribed verbatim for analyses. This allowed to trace individual cases' performace on the different measures and to compare specific protocol data to corresponding multiple-choice test data. Two graduate assistants were hired and trained by the researcher to code the protocols texts and to serve as expert judges in the analyses. Discrepancies after final checks were discussed and solved in conference.

For each text, subjects' retellings were categorized as either high, moderate or low in completeness, cohesiveness, and text understanding based on the Richness of Retelling Scale developed by Irwin and Mitchell (1983). (Interrater agreement was 88% initially, 93% in the final check.) All free-responses obtained were also analyzed for content to determine text information which was salient to subjects. The content of retellings and interview responses were then compared to verify stability of response information in the retellings. Based on the notion that questions focus subjects' responses to particular contents, the analyses were used to confirm or disconfirm specific understandings which may or may not have been apparent in the retellings.

Retelling categorization, content of retellings, and responses to interview questions were compared to responses on relevant multiple-choice test questions and to overall test scores for each subject. These analyses allowed for actual comparisons of subjects' performance on the different assessments used in the study. This specific comparison focused on information which test questions made salient but did not seem salient to subjects, and conversely, on information which seemed salient to subjects and test questions did not make salient.

For the purposes of this article, only two of the texts used in the study will be discussed in detail. The findings and issues raised in the results and discussion sections though, apply to findings in general. The specific texts discussed are described below.

Text 1. Traditional narrative, sixth grade

Title: How did Mrs. Brigley feel about owning the plates?

This narrative consisted of three paragraphs. The first paragraph dealt with Mrs. Brigley's involvement in a charity auction to help the neighborhood children buy furnishings for their clubhouse. The second paragraph is quoted below, and the third paragraph deals with Mrs. Brigley returning the porcelain plates to Dinah Hetton as a wedding gift.

Second paragraph:

> Mrs. Brigley was very surprised to see the beautiful porcelain plates that little Dinah Hetton brought. She told Dinah's mother that the porcelain was very valuable. Mrs. Hetton said it didn't matter because the plates had just been cluttering up her house. When Mrs. Brigley's bid for the plates was the highest at the auction, she was embarrassed. She thought she had bought them too reasonably. She felt guilty about even using the plates and kept them packed away. (Illinois State Board of Education & The Psychological Corporation, 1986c, p.18)

Text 2. Experimental narrative, sixth grade

Title: "The Snake Sitter"

This is a story about a girl named Linda who in order to buy her parents a wedding anniversary present, starts a pet-sitting business. One of the pets turns out to be a python named Herman. In the events of the story Herman escapes and is later found, thus resolving that problem and Linda's problem as well.

The first two paragraphs of the story read as follows:

> Linda sat on the front steps of her house, holding three dollars in her hand. She squeezed them tightly in her palm as if wishing for magic.
>
> "It's just not enough," she told Ken, her neighbor. "I wanted to buy something really special for my mom and dad's twentieth wedding anniversary. I have to have at least forty dollars so I can buy that chime clock I saw down at Gettler's Department Store. I know they'd really love it." (Illinois State Board of Education & The Psychological Corporation, 1986d, p.1)

Results

In order to illustrate findings, retellings and responses to semi-structured interview responses of four cases will be presented. Data for the four cases deal either with the traditional or the experimental sixth grade narrative. Cases and texts selected to be included here were representative of the findings of the study.

The purpose question/title of Mrs. Brigley's story suggests that the main point of the story is to understand that Mrs. Brigley felt guilty about owning the porcelain plates. However, no question in the test dealt with Mrs. Brigley's interest in helping the children (first paragraph). Three questions dealt with the plates in the following manner: the bid, where Mrs. Brigley had them stored for a long time, and Mrs. Brigley's guilt feelings. The fourth question was a main idea one and the last question was on vocabulary.

In the retellings for Mrs. Brigley's story, 7 out of 25 subjects, or 28%, clearly understood that Mrs. Brigley felt guilty about owning the porcelain plates. SH (Case 1) was one of those students as illustrated by his retelling (below). The stability of this particular understanding was corroborated by his answers to pertinent semi-structured interview questions (following the retelling protocol).

SH (Case 1):

It's about, well, the kids are going to have a charity auction for the children to raise money and they were going to get new furnishings for the clubhouse and Mrs. Brigley said she would have it in her yard and, um, this Dinah brought porcelain plates and Mrs. Brigley said they were, she was very surprised to see porcelain plates because they were very valuable... so Mrs. Brigley had the biggest bid for the plates and she was kind of embarrassed because she bid too reasonably and she felt guilty...

Q2: Who are the main characters in the story? Why are they important?

Mrs. Brigley and the children are really the main characters. Dinah Hetton comes in a little bit but not a whole lot....

Q4: What was Mrs. Brigley's problem?

Well, at the end she, um, when Dinah got married she gave it as a wedding present because she felt guilty about it and she thought she would give it somewhere else.

SH's retelling of the story was categorized as high in completeness, cohesiveness, and text understanding. His retelling demonstrated that he understood the idea that Mrs. Brigley felt guilty about owning the porcelain plates. It also demonstrated that there are other parts of the story that were important for SH such as Mrs. Brigley's involvement in the charity auction--over one-fourth of his retelling deals with this issue. From his answer to question 4, it is clear that SH understood that Mrs. Brigley felt guilty. From his answer to question 2 it is also clear that he thought Mrs. Brigley's involvement in the auction was very important.

A somewhat different example is MA (Case 2). MA's retelling was also categorized as complete, coherent, and demonstrating good text understanding. From her retelling it is not clear whether MA understood or not that Mrs. Brigley felt guilty.

MA (Case 2):

There was going to be a charity auction to raise some money for furniture for the clubhouse and um, they thought that that was the best idea that they have ever had so Mrs. Brigley or Dinah Hetton brought this expensive porcelain dishes to be auctioned off... Mrs. Brigley decides that she wanted to try buying them from the auction and she felt embarrassed because,

um, she, her bid was so reasonable and she felt embarrassed that she packed the dishes away...

Q4: What was Mrs. Brigley's problem?

Umm, she really didn't have one, I think.

Q8: What do you think Mrs. Brigley felt at the end of the story?

She felt kind of happy because she did get rid of them and that Dinah would have them.

One can conclude from the MA's retelling that she understood that Mrs. Brigley felt embarrassed because her bid was so reasonable, but it is not clear if she understood that Mrs. Brigley felt guilty. MA's rendition of Mrs. Brigley's story is representative of the retelling renditions of the majority of the sixth grade subjects for this story. The interview questions help corroborate whether MA understood or not that Mrs. Brigley felt guilty. Her answer to question 4 leads one to conclude that MA didn't understand or at least did not think Mrs. Brigley felt guilty. Her answer to question 8 doesn't help clarify. In light of the evidence one can conclude that MA thought Mrs. Brigley felt happy for any number of reasons such as getting rid of a source of embarrassment or just because Dinah was going to have the plates.

In answer to the traditional multiple-choice question, about Mrs. Brigley's guilt feelings both SH and MA selected the correct choice. The other possible choices about feelings of happiness, anger, and empathy did not fit with their text understanding as demonstrated by their retellings and their responses to semi-structured interview questions. Although MA (Case 2) selected the

correct choice it is possible it may not have exactly fit her text understanding but it was closer than the others. Also, the word guilty appeared in the text, so MA could have chosen the correct response by matching the text with the possible choices.

The next retelling, CI's (Case 3) was also rated as complete, coherent, and demonstrating good text understanding. In her retelling CI makes no mention of guilt or embarrassment on Mrs. Brigley's part.

CI (Case 3):

> The story I have just read is about a lady named Miss Brigley who wanted to put a charity auction up for some neighborhood children so that she could get some furnishings and stuff for them later on. Miss Brigley was surprised because she had seen these beautiful procelain plates... Then, Miss Brigley later on bought the plates and packed them away in a box and when Dinah, the little girl, ...

Q4: What was Mrs. Brigley's problem?

Well, um, Mrs. Brigley's problem could have been, maybe, she, help the children out because she didn't have many things to give away.

Q6: What happened in the story to help Mrs. Brigley solve her problem?

What happened was Dinah, the little girl came and gave these beautiful porcelain plates to Miss Brigley to give away in the auction, but instead Miss Brigley bought them to give to her.

Q8: What do you think Mrs. Brigley felt at the end of the story?

I think she felt very happy because she was able to help the children.

CI's retelling and answer to the interview questions clearly demonstrate that she concentrated on Mrs. Brigley's efforts to help the children furnish their clubhouse. Nowhere did she mention that Mrs. Brigley felt embarrassed or guilty. CI's answer to question 4 seems scpeculative as she probably did not believe that Mrs. Brigley had a problem. CI's response choice to the multiple-choice question about Mrs. Brigley's feelings was scored as wrong. Her choice stated that Mrs. Brigley felt happy, the closest response to her story interpretation as demonstrated by her response to question 6.

From the retellings and answers to questions, these subjects, except for Case 1, either did not understand or it is not clear if they understood the point of the story which test questions seem to address as the most important one of the story. Interestingly the raw scores of these three subjects on the passages from the traditional multiple-choice test ranged between 21 and 32 points, out of a maximum of 40. Cases 1, 2, and 3 scored 24, 32, and 21 respectively. SH (Case 1) had one of the lowest raw scores on the multiple-choice test. Recall that SH was the subject for whom it was clear that he had understood that Mrs. Brigley felt guilty about her bid for the porcelain plates. For the 5 test items relevant to the particular text discussed cases 1, 2, and 3 scored 3, 5, and 2 respectively. Figure 1 depicts sixth graders individual ratings (scores) on the three assessments for the standardized test narrative. The

scale of 5 on the vertical axis is based on the five items for the multiple-choice test, and on the 5 possible ratings on the Richness of Retelling Scale (Irwin & Mitchell, 1983).

Insert Figure 1 about here

In the retellings for text 2, The Snake Sitter, all the subjects mentioned that Linda needed money and that's why she decides to pet-sit. Twenty-one out of 25 subjects mentioned that she needed the money to buy her parents an anniversary present. The other four either did not mention it or just mentioned a present. Of 15 multiple-choice questions, one addressed Linda's desire to buy her parents a present, but did not address Linda's problem of not having money. Nine out of the 15 multiple-choice questions dealt with Herman and his escape.

Linda's problem as not having enough money was very salient in the retellings, The semi-structured interview question "What was Linda's problem?" corroborated this finding. Frequently, interviewers needed to follow-up responses with another question such as "Did Linda have any other problem?" because subjects did not mention Herman and his escape as a problem. However, 9 out of the 15 multiple-choice questions dealt with Herman and his escape.

The following retelling protocol and relevant semi-structured interview responses illustrate the saliency of Linda's problem(s) for the subjects. Case 6, VA represents one whose retelling was rated as

high in completeness, cohesiveness and text understanding.

VA, Case 4:

Linda wanted to buy her mother and her father a wedding gift, a clock, and she only had three dollars and she was holding it like she wanted magic to come and it would be more....so she ran an ad and she had people call her up, and she had three dogs, a cat, and a pet snake which was the last one to come....so they decided to give the dogs a bath, and Ken took the cat out and Linda told him cats don't get baths they wash themselves and so he put her back in and he forgot to lock it. So when the dogs saw the cat they ran off in his direction and the cat jumped out on top of the cage and one of the dogs knocked the snake cage over ...so they found all dogs and the cat and they put them all back and they couldn't find the snake and left the snake by itself and they couldn't find it and so they were thinking that they would have to buy the man another snake so then Ken was going to play a few notes on the piano and he said it sounds worser every day and so she, Linda played a note and told him he was right and so they though that it was no vibration so they thought what else could be in but a, the snake. And so Linda collected all the money....

Q4: What was Linda's problem?

She only had three dollars and she needed more money by the end of the week.

I: Did Linda have any other problems throughout the story?

Finding the snake and she thought she couldn't find the snake.

From VA's answer to question 4 and his retelling, one can observe that he percieved Linda's problem, and probably the important problem in the story, to be Linda's lack of money to buy a gift for her parents. His retelling and his answer to question 4 were typical of the 25 sixth graders in the study. VA's raw score for the multiple-choice test questions for The Snake Sitter was 58 out of 74 possible. (Item choices were scored correct when subjects made a good decision, that is, either they chose a good answer or did not choose answers which were not good. There were 74 decisions to be made and thus 74 possible points for this passage.)

As with the traditional multiple-choice test discrepancies in cases' ratings and rankings can be observed for the assessments of the experimental narrative passage. Specific discrepancies between information which is salient to readers and that which test questions make salient are probably based on different criteria. Readers seem to use text information, text characteristics, and real life problems as a source for determining what's important in a text. In contrast, test questions seem to concentrate on covering all parts of the story and on including questions of different types.

As in the case of the traditional narrative, the results for the experimental narrative point toward differences in subjects' performance when the tasks of assessment vary. It may, in addition, suggest discrepancies in the agendas of readers and of test questions.

Readers seem to establish a hierarchy of ideas based on text contents, text characteristics, and possibly on real life problems while tests seem to rely on question types and on covering different parts of the text. This may point toward differences in the processes measured as readers approach the task of reading and tests approach the task of measuring reading from different perspectives.

Discussion

The results of the present study seem to point to discrepancies in readers' text understandings, as evidenced by their retellings and interviews, and their actual test choices. The inference can be made that subjects chose a response which came closer to their text understanding or one which best fits the textual information, whether or not they thought it was the best answer to the question. Further, questions in the multiple-choice tests seem to focus on text information which is different from the information subjects report on during free-response assessment tasks.

Results were evidenced most clearly by the story about Mrs. Brigley. The title/purpose question for this passage suggests that the purpose for reading was to understand that Mrs. Brigley felt guilty about her bid for the porcelain plates at the auction. Since only 28% (7 out of 25) of the subjects in the study clearly understood that Mrs. Brigley felt guilty, one can conclude that this particular piece of information did not seem salient to most subjects. A possible explanation for this finding has to do with a discrepancy between text contents and reality; the information may be difficult to remember

because the text violates normal life expectations. Normally people are pleased when they can get good bargains rather than embarrassed or guilty about it. Another possibility is the specific context in which this particular piece of information appears in the text, which may contradict readers' predictions. First, Mrs. Hetton, who donates the plates, says they were just cluttering up her house. The next sentence says that when Mrs. Brigley's bid was the highest she felt embarrassed. Readers could possibly speculate that she was embarrassed because she bid so high for someone else's discards, but the next two sentences contradict such prediction. They state that Mrs. Brigley thought she bought them too reasonably and that she felt guilty. Most subjects answered the multiple-choice question dealing with Mrs. Brigley's guilt feelings correctly. However, in the free-response measures, that particular piece of information did not seem salient to readers and it was not clear whether they understood it or not.

The first paragraph of the story dealing with Mrs. Brigley's involvement in the charity auction was highly salient in the retellings. Approximately one-fourth of the subjects' retellings dealt with this particular issue. However, there were no multiple-choice questions about it.

A similar phenomenon was observed for other texts such as the sixth grade experimental narrative, The Snake Sitter. In The Snake Sitter, subjects focused on Linda's lack of money and everything else in the story was subordinate to that piece of

information. Such hierarchy did not seem evident in the questions included in the test. Questions seemed to attempt to address all parts of the story, as well as to include questions of different types while overlooking salient text information.

The results of the study suggest that, in general, readers' organized the information hierarchically and used that hierarchy to produce their retellings. Multiple-choice questions, on the contrary did not seem organized in the same manner.

A possible explanation for these findings is related to the specific cognitive processing required for performance on each assessment format (Davey, 1989). If this is the case, in assessing reading comprehension then, one format cannot be a substitute for the other and a variety of formats is necessary. The fact that subjects had read the texts several weeks before and re-read them for the retellings and interviews poses an interesting situation. If the multiple-choice formats and the free-response formats were measures of the same processes, then subjects should have performed better on the later tasks than on the first task and that was not the case. Subjects' performance on the initial task did not predict performance on subsequent tasks and vice versa. A plausible explanation then is that the reading processes assessed are so different that prior reading and prior performance on one task, albeit on the same text, did not have an effect on subsequent task performance. An alternative explanation is that the tasks themselves were so different from one another that their format affected performance although the processes

measured were the same. This argument would be incompatible with the established trait theory. According to the theory, a good reader ought to perform equally well on all types of assessment formats assessing the same trait. Good or poor reading behaviors are considered to be all reader related factors. However, the evidence presented indicates that reading ability may be affected by an array of factors such as text related factors and task related factors, plus reader related factors.

Consistent with the argument about the difference in tasks would be the notion that verbally delivered formats direct readers more clearly than written formats. In verbal formats examiners can always expand and explain differently while in written formats readers need to rely on their reading proficiency. This argument requires serious consideration. Reading and listening require different perceptual processings. However, if verbal directions and questions have generalized advantages over written ones, then all subjects should have performaned higher on the verbally administered tasks and that was not the case. Along the same lines would be the argument about subjects' ability to verbalize their ideas. Again, the processing would be different than that required to read and select responses, but in light of the evidence it would be difficult to argue that one would present definite advantages over the other.

Another possible explanation for the difference in results between response formats is that subjects just did not remember enough of the texts from the first reading to affect results. Subjects

indicated that they remembered taking the tests and reading the passages but they may not have remembered enough information to affect performance.

In sum, the results point to discrepancies in the use of text information by subjects and by test questions. Salient information in third and sixth graders verbal responses seem to differ from information made salient by test questions. This was evidenced by the content of subjects' responses in the different measures used. Varying subjects' ratings and rankings different measures give support to the notion that multiple-choice assessments and free-response assessments are measuring different processes or different components of the same process. These results suggest that text content and assessment format may play an important role in assessing the reading process.

Implications

This study has raised issues about reading assessment which are relevant for test development. Although somewhat speculative at this point, these implications may be worth considering for further study in reading assessment. First, and of immediate application, would be in improving existing tests. Second, would be to change, or at least modify, the current assessment paradigm in reading.

One way in which currently used tests could possibly be improved is by the use of free-response formats as part of the test development process. Retellings and responses to interview questions can be useful in selecting test passages and in developing test

questions and their respective choices in multiple-choice tests. Such a procedure should help detect texts which are confusing incompatible with readers' beliefs. Free-response formats can also yield information about text content which is salient to readers. In addition, readers' text understandings and misunderstandings could be analyzed to develop distractors which are in various degrees of conflict with those understandings.

A point to consider is that scoring free-response formats need not be complicated or time-consuming. Wholistic scorings such as the ones used in this study are simple to use once scorers are trained and well versed on their use. Granted, these will take longer to perform than machine-scoring, but the long term gains in test development may more than compensate for the additional time spent.

Along with the use of wholistic scorings and free-response formats is the possibility of weighing test questions according to various levels of understanding. A question that addresses the thesis of an exposition or the overriding problem of a narrative, as evidenced by free-response formats, would carry more score weight than a subordinate problem or a supporting detail. This notion goes beyond the present taxonomy of questions into inferential, cause and effect, literal, and so on. It deals with text characteristics in terms of organization of content as evidenced by readers' renditions of understanding. Questions can be classified according to whether they address information considered salient, superordinate, subordinate, or as the task or text requires as evidenced by readers.

In case of divergent responses, the specific text interpretation may be analyzed further for plausibility of interpretation. Thus, infrequent but plausible interpretations can be scored in such a way that readers' use of legitimate comprehension strategies are acknowledged. In addition, a measure of cohesiveness should be introduced to determine the degree of cohesiveness of understanding based on the composite of choice selections. Such measure must be based on whether the composite of subjects' choice selections represent a cohesive understanding of a text.

In essence, much research in testing practices as they relate to reading theory is needed. Research on alternative formats, content of response, quality of response, saliency of textual information, and cohesiveness of understanding, are but a few examples of this need. As the existing gap between testing and reading theory narrows, instructional practice, which seems to be trapped in the existing dichotomy, will benefit.

REFERENCES

ANDERSON, R. C. (1977). The notion of schemata and the educational enterprise: General discussion of the conference. In Anderson, R. C., Spiro, R. J., & Montague, W. E. (Eds.), _shooling and the acquisition of knowledge_ (pp. 415-431). Hillsdale, NJ: Erlbaum.

ANDERSON, R. C. & ORTONY, A. (1975). On putting apples into bottles: A problem of polysemy. _Cognitive Psychology_, _7_, 167-180.

BARTLETT, F. C. (1932). _Remembering: A study in experimental and social psychology_. London: Cambridge University Press.

BOCK, J. K. & BREWER, W. F. (1974). Reconstructive recall in sentences with alternative surface structures. _Journal of Experimental Psychology_, _103_, 837-843.

COLE, N. S. (1984). Testing and the "crisis" in education. _Educational Measurement_, _3_(3), 4-8.

DAVEY, B. (1989). Assessing comprehension: Selected interactions of task and reader. _The Reading Teacher_, _42_, 694-697.

FARR, R., & CAREY, R. F. (1986). _Reading, What can be measured?_ Newark, DE: IRA.

FREDERIKSEN, N. (1984). The real test bias: Influences of testing on teaching and learning. _American Psychologist_, _39_, 193-202.

ILLINOIS STATE BOARD OF EDUCATION & THE PSYCHOLOGICAL CORPORATION (1986a). _The Illinois Inventory of Educational Progress, Inventory: 1, Form 311_. Springfield, IL: Harcourt, Brace, Jovanovich.

ILLINOIS STATE BOARD OF EDUCATION & THE PSYCHOLOGICAL CORPORATION (1986b). The Illinois Inventory of Educational Progress, Inventory: 3, Form 333. Springfield, IL: Harcourt, Brace, Jovanovich.

ILLINOIS STATE BOARD OF EDUCATION & THE PSYCHOLOGICAL CORPORATION (1986c). The Illinois Inventory of Educational Progress, Inventory: 1, Form 611. Springfield, IL: Harcourt, Brace, Jovanovich.

ILLINOIS STATE BOARD OF EDUCATION & THE PSYCHOLOGICAL CORPORATION (1986d). The Illinois Inventory of Educational Progress, Inventory: 3, Form 633. Springfield, IL: Harcourt, Brace, Jovanovich.

IRWIN, P. A., & MITCHELL, J. N. (1983). A procedure for assessing the richness of retellings. Journal of Reading, 26, 391-396.

JOHNSTON, P. H. (1983). Reading comprehension assessment: A cognitive basis. Newark, DE: IRA.

JOHNSTON, P. H. (1984). Assessment in reading. In P. D. Pearson (Ed.), Handbook of reading research (pp. 147-182). NY: Longman.

MADAUS, G. F. (1983). Minimum competency testing for certification: The evolution and evaluation of test validity. In G. F. Madaus, (Ed.), The courts, validity, and minimum competency testing (pp. 21-61). Boston: Kluwer-Nijhoff.

MESSICK, S. (1975). The standard problem: Meaning and values in measurement and evaluation. American Psychologist, 30, 955-966.

MESSICK, S. (1981). Constructs and their vicissitudes in educational and psychological measurement. Psychological Bulletin, 89, 575-588.

NORMAN, D. A., & RUMELHART, D. E. (1975). Memory and knowledge. In D. A. Norman & D. E. Rumelhart (Eds.), Explorations in cognition. San Francisco: Freeman.

Diego, CA.

PRESCOTT, W., BALOW, I. H., HOGAN, P. P., & FARR, R. C. (1984). Metropolitan Achievement Test (6th edition). San Antonio, TX: The Psychological Corporation.

PEARSON, P. D., & JOHNSON, D. (1978). Teaching reading comprehension. NY: Holt.

RUMELHART, D. E., & ORTONY, A. (1977). The representation of knowledge in memory. In R. Anderson, R. J. Spiro, & W. Montague (Eds.),Schooling and the acquisition of knowledge (pp. 99-130). Hillsdale, NJ: Erlbaum.

SCHALLERT, D., ULLERICK, S., & TIERNEY, R. J. 1984). Evolving a description of text through mapping. In C. D. Holley & D. Dansereau (Eds.), Special learners: Techniques, application and related issues (pp. 255-274). NY: Academic Press.

SMITH, N. B. (1965). American reading instruction. Newark, DE: International Reading Association.

TAYLOR, B. M. (1984). The search for a meaningful approach to assessing comprehension of expository text. In Niles, J. A. & Harris, L. A. (Eds.), Thirty-third Yearbook of the National Reading Conference (pp. 257-263). Rochester, NY: NRC.

THORNDIKE, E. L. (1917). Reading as reasoning: A study of mistakes in paragraph reading. Journal of Educational Research, 8, 823-832.

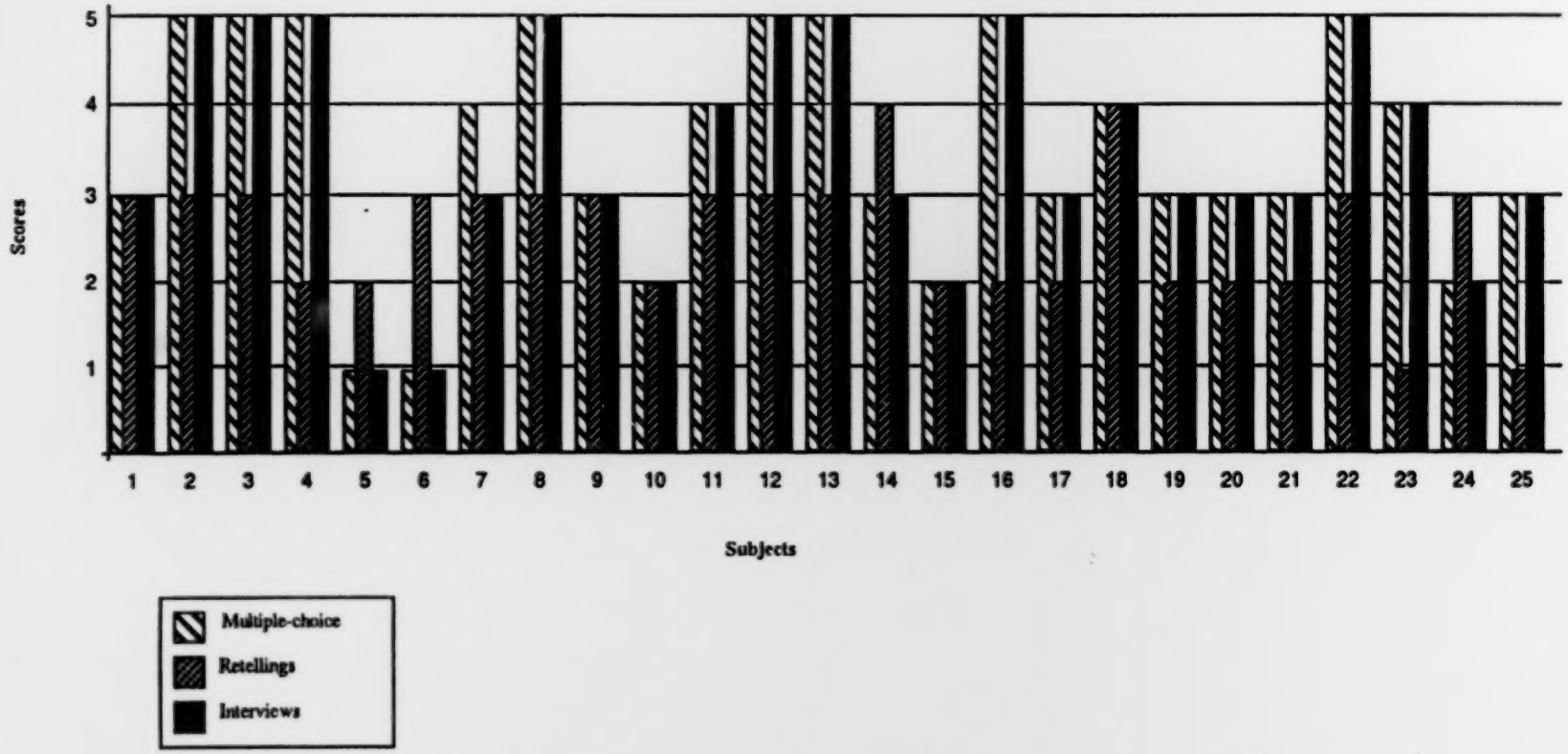
TIERNEY, R. J., & McGINLEY, W. (1987, April). Serious flaws in written literacy assessment Paper presented at the annual Meeting of the American Educational Research Association, Washington, D.C.

VALENCIA, S., & PEARSON, P. D. (1987). Reading assessment: Time for a change. The Reading Teacher, 40, 726-732.

WARREN, W. H., NICHOLAS, D. N., & TRABASSO, T. (1979). Event chains and inferences in understanding narratives. In R. O. Freedle (Ed.), New directions in discourse processing: Advances in discourse processes (Vol. 2, pp. 23-52). Norwood, NJ: Ablex.

WIXSON, K. K., PETERS, C. W., WEBER, E. M., ROEBER, E. D. (1987). New directions in statewide assessment. The Reading Teacher, 40, 749-754.

Figure 1. Sixth graders ratings (scores) on the assessments for the standardized test narrative



A comparison of sixth graders' scores by standardized passage measures